We would like to thank the reviewer for the time spent on the review, prior and during the discussion phase.
Our answers appear in blue below.

Anonymous Referee #3

The topic of this manuscript is operational satellite measurements around the 183 GHz water vapour transition, including how well these measurements can be reproduced by radiative transfer simulations. Such measurements provide today a large fraction of the information on water vapour in weather forecasts, and the same measurements increase in importance for climate monitoring. Accordingly, the manuscript deals with an important issue and fits AMT well.

I agree with two other referees, both on a general level and in points raised. The manuscript should be published, after some minor changes. Also I would like to see a conclusion section. What are the most plausible reasons to the discrepancies discussed? That is not totally clear after reading the manuscript. As referee 1 I found the text unclear in some places, mainly due to complex sentences and unlucky choices of words, but I leave this as a general remark.

My main demand is to specify best estimates of relevant spectroscopic variables (line parameters + continua). On lines 196-197 some uncertainty limits are given (which I personally find surprisingly small), but around what values do these values apply? Further, are the assumed best estimates and the error estimates on lines 196-197 consistent with the findings from field measurements discussed around line 216 and 328? And are assumed best estimates used when producing Figure 1 and 2? A table on spectroscopic parameters used in main models (such as MPM, RTTOV and CRTM) would be very useful.
This is a fair remark and we have thus added a table (namely Table 3 “Spectroscopic parameters of the 183-GHz line at 296 K”, see below) in order to compare the spectroscopic variables from lab measurements, the HITRAN2012 list and two line-by-line models, MonoRTM and ARTS.
RTTOV and CRTM are not line-by-line models, and depend on the line-by-line models that were used to train them. However, in the Clain et al. (2015) paper, RTTOV, ARTS and MonoRTM have been compared and show the same channel-dependent bias in model vs SAPHIR comparisons, although they have a different line shape parametrization.

Also, it is very important to underline that while HITRAN is a very popular and permanently updated database, its values have associated uncertainties, which are discussed during HITRAN committee meetings. This statement is now included in the paper to make the reader aware that HITRAN shouldn’t be considered as a unique source for spectroscopic parameters.

Concerning the continua, it is more complex. Indeed there is no universal model of the continua yet and they are defined from the difference between observed spectra and those calculated by some line-by-line model. So the “best estimates of the relevant parameters” of the continua is line-shape model dependent. Giving a number may cause potential problems of improper usage. We feel that the question of the continua is sufficiently covered by Figure 3 and the related text. Interested readers can find corresponding parameters in given references.

Source	Spectroscopic parameters and units. ** Integrated intensity is given taking into account natural abundance of $H_2^{16}O$ isotopologue (99.73(3)%) in accordance with HITRAN's convention.*								
	Line central frequency (kHz)	Integrated intensity (10^{-25} cm/mol)	Low level energy (cm^{-1})	Air broadening (MHz/Torr)	Self broadening (MHz|Torr)	Air shifting (MHz|Torr)	Self shifting (MHz|Torr)	Broadening T-exponent	Shifting T-exponent
Tretyakov (2016) - lab. measurements -	183 310 087(1)	785(8)*	136.163927(1)	3.926(20)	19.7(5)	-0.096(10)	+0.23(3)	0.74(3)	0.76(5)
HITRAN 2012 - line list - *uncertainties*	183 310 11	778.5 *5-10%*	136.1639	3.932 *2-5%*	20.5 *2-5%*	-0.107 *0.03-0.3*	not listed	0.68 *2-5%*	not listed
MonoRTM v4.2 - model -	183 310 11	769.1	136.1639	3.932(100)	17.7	-0.106	not accounted for	0.77	not accounted for
ARTS v2.3.3 ARTS v2.3.4 - model -	183 310 11 183 310 11	758.1 778.5	136.1639 136.1639	3.779 3.933	20.107 21.530	not listed -0.106	not listed not listed	0.64 (air) 0.85 (self) 0.77 (both)	not listed

Table 3: Spectroscopic parameters of the 183-GHz line at 296 K assuming a Voigt line shape. RTS version v2.3.3 followed Rosenkranz (1998) specifications and was used in Clain et al. (2015). ARTS v2.3.4 is the current set-up and mixes Payne et al. (2008) parameters with HITRAN data for parameters not covered there.

And even better would be if the values used for AURA MLS could be included. Are the MLS values inside the specified uncertainty range?

We did not MLS in the paper because this sounder is a passive limb sounder and is primarily aimed at measuring water vapor in the stratosphere, so both a very different geometry and aslo region to the biases we have been investigating.

However, a link to MLS is now done in the recommendation section 3 since stratospheric water vapor estimated from MLS is a considered as reference dataset.

The authors could also help the reader a bit more. The usage of different units is probably unavoidable, but it would be useful to provide some rough mapping between units. For example, can a rough scaling between e.g. 5% error of in-situ data and a simulated brightness temperature be given? Another example is the sentence on lines 333-335. It would be easier for the reader to relate the results to the bias seen in Fig1. Are the results based on ATOMMS consistent the sign of the bias in Fig 1, or is it reversed?

The ATOMMS results are entirely in sync with figure 1, where we see that the modeled BT is higher than the observed one. This is worth highlighting the consistency of the two, so we have added the following sentence in the ATOMMS discussion:

"Viewed from space (…). This result is consistent with the discrepancies between the satellite-based measurements and the modeled one under scrutiny here. "

Smaller comments:

Line 47: I don't see the point in comparing systematic errors with the noise. The systematic error should (and can) be small even for a noisy instrument.

For real-time applications, if a systematic error is small compared to noise, then its correction using standard bias correction is likely to be good enough. If it is large compared to the noise then this raises a question.

It is true that for applications, such as climate-related studies, the biases are important even if they are small compared to the noise: the random-character of noise makes it less important for such study.

Anyhow, we do not wish to include such discussion in the paper, because we think it is out of scope.

Line 65: "snow of", should "snow and"? (In any case, would be good to clarify if falling or surface snow is discussed).
Indeed we recognize that the sentence was not clear: we meant thick ice clouds.
We re-wrote the sentence and it now reads
“While large differences can be seen in areas with a high occurrence of optically thick (at 183GHz) ice clouds, there is a clear background difference, between 60°N and 60°S, that increases towards the line wings and that cannot be related to a particular region or atmospheric state”

Line 116: Can a reference from 2007 be valid for most recent campaign?
We do not understand why a study performed in 2007 would not be relevant.

Line 118-119: Unclear. What is meant with "considerably small"? That is, what is compared here, different atmospheric layers?
This part discusses the impact of a fractional modification of the measured atmospheric RH, induced by a sensor measuring capacity, on the measured BT. We have clarified this and it now reads:
“Also, a given fractional error in humidity, similar to the accuracy of the sensor, causes a considerably smaller fractional change of BTs.”

Also, it should be “considerably smaller” rather than “small”. We corrected it.

Line 123: It must be motivated why GNSS is included. The unclear link to the topic of the manuscript is also reflected in the last sentence of the paragraph ("might"!). And clarify that the topic is ground-based GNSS.
We included the term “Ground-based” at the beginning of the section.
Indeed, the justification of why we mention GNSS was lacking. Therefore we have added the following sentence at this end of the paragraph:
“Indeed, since the wings of the 183.31GHz line are mainly sensitive to the lower tropospheric humidity, and the PW is dominated by the humidity content of the lower troposphere. The analysis of the GNSS-estimated PW provides hints at the potential for a contribution of the PW to the overall 183GHz biases.”

Line 175: To my best knowledge, ARTS can make use of several water vapour continua models.
Indeed, ARTS is a flexible line-by-line model and can use several different continuum models, but MT_CKD is nowadays the most popular one within the radiometric community. There is more discussion on continua further down in the manuscript, the point we wanted to make here simply was that although the same model is used for IR and MW, this does not mean that biases have to be similar in the two different spectral regions.
The text was rephrased to make this point clearer:
- Line 175: “These three RTMs rely on the currently most widely accepted model MT_CKD (MlawerTobin_CloughKneisysDavies, [Mlawer et al., 2012]) for the parametrization of the absorption due to the water vapor continuum.”
- Line 187: “This does not necessarily mean that the same mechanisms must be responsible for the biases observed in the MW radiances, even though the continuum model MT_CKD covers both spectral regions. The MT_CKD model is semi-empirical and thus could have different biases in the IR and the MW spectral regions.”
Moreover, we added a line for ARTS in the new Table 3.

Line 342: I don’t see the logic in "Accordingly" here.
We have replaced the term “Accordingly” by “This suggests that”: indeed the comparison is only made between two systems, which can only provide an assumption on the interpretation.

Text of figure 1: Is that really the name of a single campaign? That seems to be two or three campaigns. And why defining acronyms here?
The CINDY/DYNAMO/AMIE international field experiment gathered three initiatives around the study of the MJO in the Indian Ocean, during the same period.
We moved the definition of the acronyms of this campaign in the main body of the manuscript, around line 129, when we first mention it.
Along this comment, we have added an appendix listing the acronyms.